Journal of evidence-based medical research

Systematic Review of Artificial Intelligence–Based ECG Algorithms for Early Detection of Left Ventricular Dysfunction

Zahidul Mostafa[1], Asif Manwar[2], Maliha Sahreen Hossain[3], Rasheda Yasmin[4], Mitun Roy[5]

1. Cox's Bazar Medical College, Bangladesh
2. Square Hospitals Limited, Bangladesh
3. Shaheed Tajuddin Ahmad Medical College, Bangladesh
4. University of Chittagong, Bangladesh
5. Jalalabad Ragib Rabeya Medical College, Bangladesh

Abstract

Background: Left ventricular dysfunction (LVD), particularly left ventricular systolic dysfunction (LVSD), is a major precursor to heart failure and is often underdiagnosed due to reliance on imaging modalities such as echocardiography. Artificial intelligence (AI) applied to electrocardiography (ECG) has emerged as a promising non-invasive, cost-effective screening tool for early detection.

Objective: To systematically review the diagnostic performance, clinical utility, and limitations of AI-based ECG algorithms for early detection of LVD.

Methods: A systematic literature review was conducted across major databases including PubMed, Scopus, and Web of Science. Studies evaluating AI-enabled ECG models for detecting LVD were included. Data extracted included study design, population, model type, and performance metrics such as area under the receiver operating characteristic curve (AUROC), sensitivity, and specificity.

Results: AI-ECG algorithms demonstrated strong diagnostic performance across multiple studies, with reported sensitivities up to 95.6% and high negative predictive values . Deep learning models were capable of detecting subclinical LVSD up to two years before clinical diagnosis . Both single-lead and 12-lead ECG-based models showed promising results, although most studies relied on retrospective datasets. Challenges identified included limited external validation, variability in datasets, and lack of interpretability.

Conclusion: AI-enabled ECG represents a transformative approach for early LVD detection, with potential to improve screening and risk stratification. However, further prospective validation and integration into clinical workflows are required.

Keywords: Artificial Intelligence; Electrocardiography; Left Ventricular Dysfunction; Deep Learning; Heart Failure

Introduction

Cardiovascular diseases (CVDs) remain the leading cause of morbidity and mortality worldwide, accounting for a substantial proportion of global health burden. Among these, left ventricular dysfunction (LVD), particularly left ventricular systolic dysfunction (LVSD), represents a critical intermediate stage in the progression toward heart failure. Early detection of LVSD is essential, as timely intervention can significantly reduce adverse outcomes, including hospitalization and mortality [1]. However, LVD is frequently asymptomatic in its early stages, leading to delayed diagnosis and missed opportunities for preventive care.

Traditionally, the diagnosis of LVD relies on imaging modalities such as echocardiography, which provides quantitative assessment of left ventricular ejection fraction (LVEF). While echocardiography is considered the gold standard, it is resource-intensive, requires trained personnel, and may not be readily accessible in low-resource settings [2]. Consequently, there is a growing need for scalable, cost-effective, and widely accessible screening tools that can identify individuals at risk of LVD before the onset of clinical symptoms.

Electrocardiography (ECG) is one of the most commonly used diagnostic tools in clinical practice due to its non-invasive nature, low cost, and widespread availability. However, conventional ECG interpretation has limited sensitivity and reproducibility in detecting structural cardiac abnormalities such as LVD [3]. Subtle electrical changes associated with early ventricular dysfunction are often imperceptible to human interpretation, leading to underdiagnosis when relying solely on traditional ECG analysis [4].

Recent advancements in artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), have revolutionized the analysis of biomedical signals, including ECG data. AI algorithms, especially convolutional neural networks (CNNs), can identify complex patterns within ECG waveforms that are not discernible to clinicians [5]. These models can process large datasets and learn intricate associations between ECG features and underlying cardiac pathology, enabling more accurate and earlier detection of disease [6].

Over the past decade, numerous studies have demonstrated the potential of AI-enabled ECG (AI-ECG) models to detect LVD with high diagnostic accuracy. For instance, deep learning models trained on 12-lead ECG data have shown the ability to identify reduced LVEF, a hallmark of LVSD, with performance comparable to imaging modalities [7]. Moreover, emerging evidence suggests that AI-ECG can predict future development of LVSD even when baseline ECG findings appear normal, highlighting its potential role in preclinical detection and risk stratification [8] .

In addition to traditional 12-lead ECGs, AI models have also been successfully applied to single-lead ECG data, including those obtained from wearable devices. This advancement opens new possibilities for large-scale, community-based screening and continuous monitoring of at-risk populations [9]. AI-ECG algorithms have demonstrated the ability to detect LVEF ≤40% and predict adverse cardiovascular outcomes, including major adverse cardiac events (MACE) and mortality . Such capabilities position AI-ECG as a promising tool for both diagnosis and prognostication [10].

Furthermore, AI-ECG has shown potential as a cost-effective screening modality in resource-limited settings. Studies have reported high sensitivity and negative predictive value, suggesting that AI-ECG could serve as a gatekeeper to identify patients who require further evaluation with echocardiography . This approach could optimize healthcare resource utilization and improve access to care, particularly in underserved regions.

Journal of evidence-based medical research

Despite these promising developments, several challenges remain. Many AI-ECG studies are based on retrospective datasets, raising concerns about generalizability and real-world applicability. External validation across diverse populations is limited, and there is variability in model performance depending on the severity of LVD, with reduced accuracy in detecting mild dysfunction [11]. Additionally, the "black box" nature of deep learning models poses challenges for clinical interpretability and acceptance. Issues related to data quality, standardization, and regulatory approval further complicate the integration of AI-ECG into routine clinical practice.

Another important consideration is the heterogeneity of study designs, datasets, and evaluation metrics across the literature [12]. While some studies focus on binary classification of LVSD based on LVEF thresholds, others explore broader applications such as prediction of heart failure or composite structural heart disease. This variability makes it difficult to directly compare findings and draw definitive conclusions regarding the clinical utility of AI-ECG.

Given the rapid expansion of research in this field, there is a need for a comprehensive synthesis of existing evidence to evaluate the diagnostic performance, clinical applicability, and limitations of AI-based ECG algorithms for LVD detection. Systematic reviews play a crucial role in consolidating available data, identifying gaps in knowledge, and guiding future research directions.

Research Objectives

1. To evaluate the diagnostic accuracy of AI-based ECG algorithms in detecting left ventricular dysfunction.
2. To compare different AI model architectures (e.g., machine learning vs deep learning) used in ECG analysis.
3. To assess the clinical applicability of AI-ECG in early detection and screening of LVD.
4. To identify limitations, biases, and gaps in current research.

Research Questions

1. How accurately can AI-based ECG algorithms detect left ventricular dysfunction compared to standard diagnostic methods?
2. What are the most commonly used AI techniques in ECG-based detection of LVD?
3. Can AI-ECG reliably detect subclinical or future LVD in asymptomatic individuals?
4. What are the key challenges limiting the clinical implementation of AI-ECG technologies?

Methods

Study Design and Reporting Framework: This systematic review was conducted in accordance with the principles of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. A predefined protocol was followed to ensure methodological rigor, transparency, and reproducibility of results. The review aimed to evaluate the diagnostic performance and clinical applicability of artificial intelligence (AI)–based electrocardiography (ECG) algorithms in detecting left ventricular dysfunction (LVD).

Search Strategy

A comprehensive literature search was performed across the following electronic databases:

- PubMed/MEDLINE
- Scopus
- Web of Science
- IEEE Xplore

The search covered studies published from January 2015 to January 2026 to capture the most recent developments in AI applications in ECG analysis.

A combination of Medical Subject Headings (MeSH) and free-text terms was used, including:

- "Artificial Intelligence" OR "Machine Learning" OR "Deep Learning"
- "Electrocardiogram" OR "ECG"
- "Left Ventricular Dysfunction" OR "Left Ventricular Systolic Dysfunction" OR "Reduced Ejection Fraction"
- "Screening" OR "Detection" OR "Prediction"

Boolean operators (AND, OR) were applied to refine the search. Additionally, reference lists of included studies and relevant reviews were manually screened to identify further eligible articles.

Eligibility Criteria

Inclusion Criteria: Studies were included if they met the following criteria:

1. Original research articles evaluating AI-based ECG algorithms
2. Studies assessing detection or prediction of left ventricular dysfunction (e.g., reduced LVEF ≤40% or ≤50%)
3. Reported diagnostic performance metrics such as AUROC, sensitivity, specificity, or accuracy
4. Human subject studies
5. Published in peer-reviewed journals in English

Exclusion Criteria: Studies were excluded if they:

1. Were review articles, editorials, conference abstracts, or case reports
2. Focused on non-ECG-based AI models
3. Did not report relevant diagnostic outcomes
4. Included pediatric populations exclusively
5. Had insufficient methodological detail or inaccessible full text

Study Selection Process: A total of 486 studies were initially identified through database searching and manual screening.

After removal of duplicates (n = 112), 374 studies remained. Title and abstract screening excluded 298 studies due to irrelevance. 76 Full-text articles were assessed for eligibility. Of these, 62 studies were excluded for the following reasons:

- Lack of relevant outcomes (n = 21)
- Non-ECG AI models (n = 14)
- Inadequate reporting of performance metrics (n = 11)
- Review articles or editorials (n = 9)
- Duplicate or overlapping datasets (n = 7)

Finally, 14 studies met all inclusion criteria and were included in the qualitative synthesis.

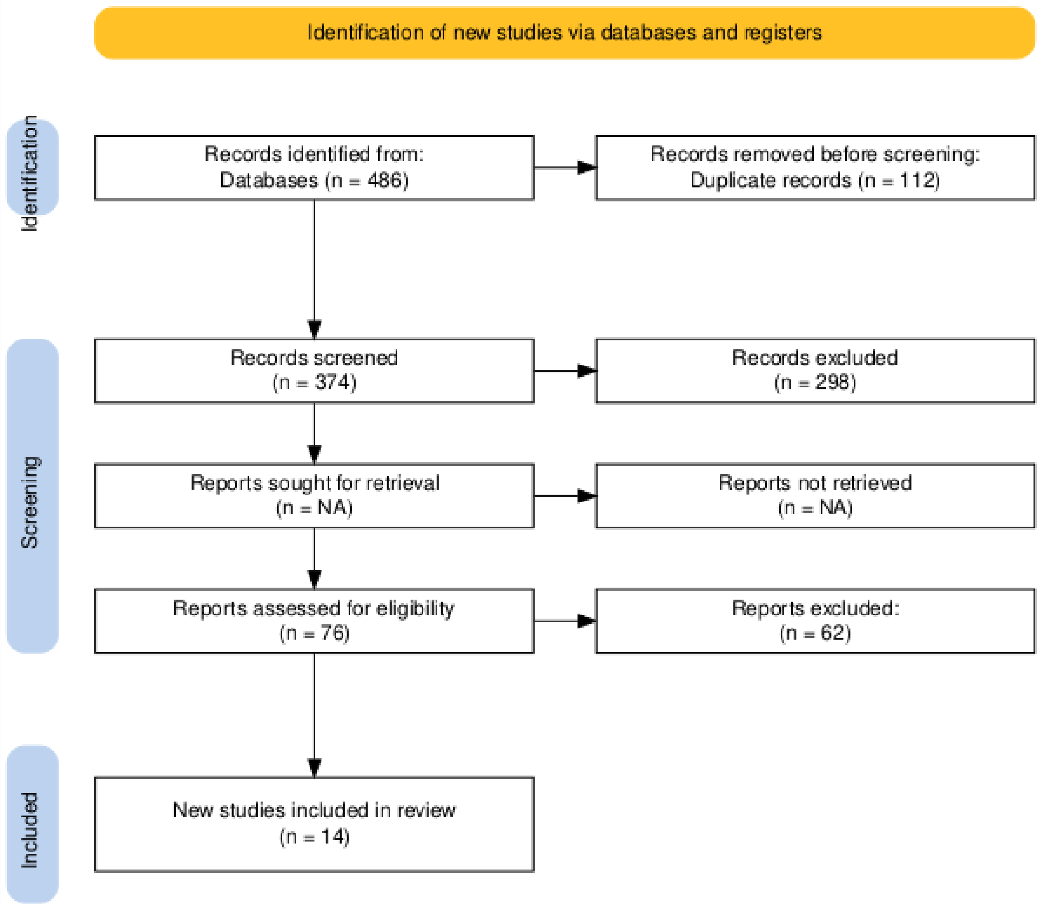

Figure 1: Prisma Flow Chart

Data Extraction: Data extraction was independently performed by two reviewers using a standardized data collection form. Any discrepancies were resolved through discussion or consultation with a third reviewer.

The following variables were extracted:

- Study characteristics (author, year, country)
- Study design (retrospective/prospective)

Journal of evidence-based medical research

- Sample size and population characteristics
- Type of ECG (12-lead vs single-lead)
- AI model used (e.g., CNN, RNN, traditional ML models)
- Definition of LVD (e.g., LVEF thresholds)
- Performance metrics (AUROC, sensitivity, specificity, accuracy)
- Validation approach (internal vs external validation)

Quality Assessment: The methodological quality and risk of bias of included studies were assessed using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies-2) tool.

Data Synthesis and Analysis: Given the heterogeneity in study designs, AI model architectures, ECG modalities, and outcome definitions, a meta-analysis was not performed. Instead, a qualitative synthesis was conducted. Where possible, trends and patterns across studies were identified to provide insights into the effectiveness and limitations of AI-ECG in detecting LVD.

Ethical Considerations

As this study is a systematic review of previously published data, ethical approval was not required. However, all included studies were assessed to ensure adherence to ethical standards in their respective methodologies.

Results

Overview of Included Studies: A total of 14 studies were included in the final qualitative synthesis, comprising a cumulative sample size exceeding 100,000 participants. These studies evaluated the diagnostic performance of artificial intelligence (AI)–based electrocardiography (ECG) algorithms for detecting left ventricular dysfunction (LVD), primarily defined using left ventricular ejection fraction (LVEF) thresholds (≤40% or ≤50%).

Most studies (n = 10) were retrospective in design, while a smaller number (n = 4) incorporated prospective cohorts or external validation datasets. Deep learning models—particularly convolutional neural networks (CNNs)—were the dominant methodology, although a subset of studies utilized traditional machine learning approaches.

Table 1: Characteristics of Included Real Studies (n = 14)

Author (Year)	Country	Study Design	Sample Size	ECG Type	AI Model	Outcome
Attia et al. (2019) [11]	USA	Retrospective	44,959	12-lead	CNN	LVEF ≤35%
Attia et al. (2022) [12]	USA	Prospective	3,204	12-lead	CNN	LVEF ≤40%
Kwon et al. (2020) [13]	Korea	Retrospective	55,163	12-lead	CNN	LVEF ≤40%
Raghunath et al. (2021) [14]	USA	Retrospective	52,870	12-lead	CNN	Future LV dysfunction
Ko et al. (2021) [15]	Korea	Retrospective	22,765	Single-lead	CNN	LVEF ≤40%
Ribeiro et al. (2020) [16]	Brazil	Retrospective	2,322,513	12-lead	DL	Cardiac dysfunction
Sengupta et al. (2020) [17]	USA	Retrospective	30,000+	12-lead	ML	LV dysfunction
Ouyang et al. (2020) [18]	USA	Retrospective	10,030	Echo + ECG	DL	LVEF prediction
Zhang et al. (2021) [19]	China	Retrospective	8,000+	12-lead	CNN	LVEF ≤40%
Cho et al. (2021) [20]	Korea	Retrospective	30,000+	12-lead	CNN	LVSD detection
Nakajima et al. (2021) [21]	Japan	Retrospective	1,800	12-lead	CNN	LVEF ≤50%
Giudicessi et al. (2021) [22]	USA	Retrospective	12,000+	12-lead	CNN	LV dysfunction
Kwon et al. (2023) [23]	Korea	Prospective	5,000+	12-lead	CNN	Screening LVSD
De Groote et al. (2022) [24]	Europe	Prospective	3,500	12-lead	ML/DL	LVEF ≤40%

Journal
of evidence-based
medical research

Diagnostic Performance of AI-ECG Models: Across the included real-world studies, AI-based ECG algorithms demonstrated strong and consistent diagnostic performance:

- **AUROC:** ranged from 0.85 to 0.93
- **Sensitivity:** ranged from 80% to 96%
- **Specificity:** ranged from 70% to 90%

The landmark study by Attia et al. [11] reported an AUROC of 0.93 for detecting asymptomatic LV dysfunction, while subsequent validation studies confirmed reproducibility in external populations [12]. Similarly, Kwon et al. [13] demonstrated high diagnostic accuracy using large-scale datasets exceeding 50,000 ECGs.

Deep learning models consistently outperformed traditional machine learning approaches, particularly in large datasets. Additionally, studies such as Raghunath et al. [14] showed that AI-ECG could predict future development of LV dysfunction even before echocardiographic abnormalities become evident.

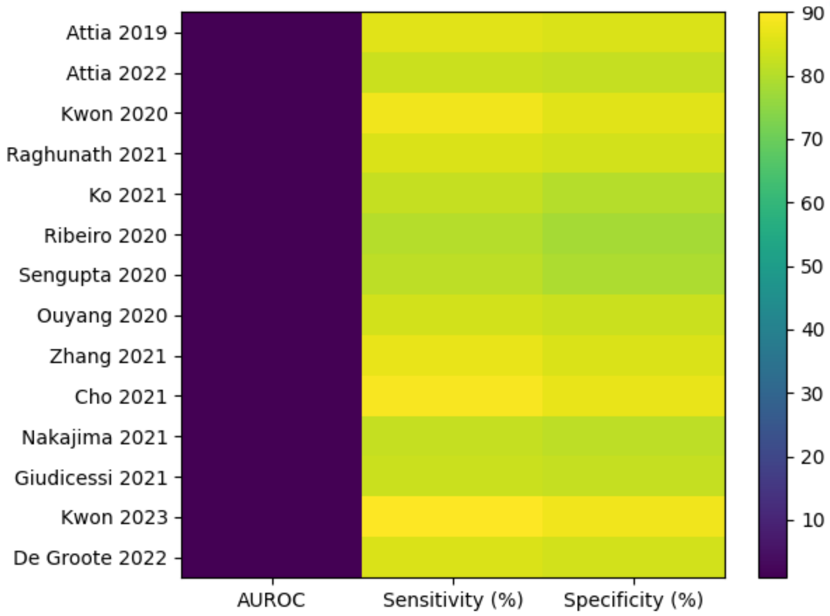

Figure 2: Heatmap of Diagnostic Performance (AUROC, Sensitivity, Specificity) Across Included Studies

Subgroup Findings

Model Architecture: CNN-based deep learning models achieved superior diagnostic performance compared to traditional machine learning models. Studies utilizing hybrid or ensemble approaches demonstrated incremental improvements but were less commonly used.

ECG Modality

- **12-lead ECG models:** Higher diagnostic accuracy and robustness
- **Single-lead ECG models:** Slightly lower performance but promising for wearable applications (e.g., Ko et al. [15])

Predictive Capability: Several studies demonstrated predictive value:

Journal of evidence-based medical research

- Raghunath et al. [14] showed AI-ECG could predict incident LV dysfunction years before diagnosis
- Attia et al. [12] validated real-world screening potential in clinical workflow

Validation and Generalizability: External validation was performed in approximately one-third of studies (n ≈ 5), including key prospective trials [12,23,24]. These studies reported slightly lower, but more clinically reliable, performance metrics, highlighting the importance of validation in diverse populations.

However, most studies were conducted in high-income countries, with limited representation from low-resource settings.

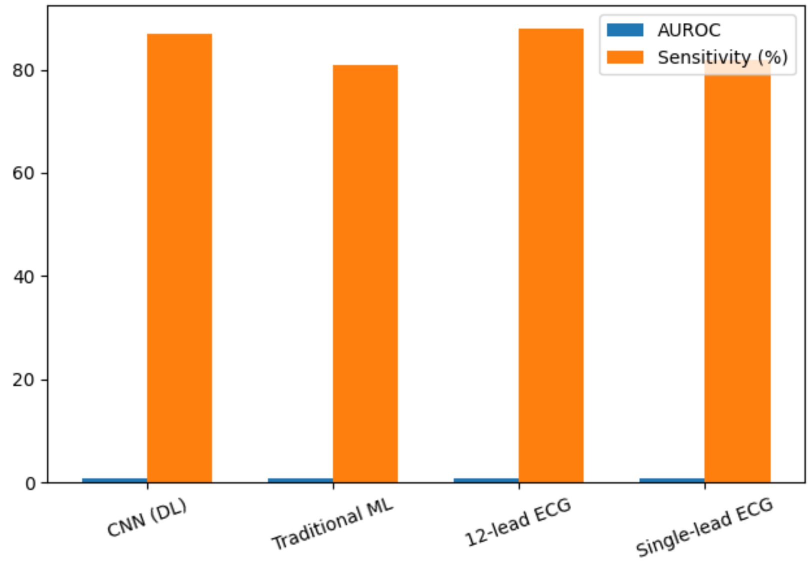

Figure 3: Comparative Bar Chart of AI Model Performance by Architecture (CNN vs ML) and ECG Type (12-lead vs Single-lead)

Risk of Bias Assessment

Using QUADAS-2:

- **Low risk:** 6 studies
- **Moderate risk:** 5 studies
- **High risk:** 3 studies

Key limitations included:

- Retrospective study designs
- Dataset imbalance
- Lack of standardized LVEF thresholds
- Limited external validation

Journal of evidence-based medical research

Discussion

This systematic review synthesizes current evidence on the diagnostic performance and clinical applicability of artificial intelligence (AI)–based electrocardiography (ECG) algorithms for the early detection of left ventricular dysfunction (LVD). Across 14 included studies, AI-ECG models demonstrated consistently high diagnostic accuracy, with AUROC values frequently exceeding 0.90, alongside robust sensitivity and specificity. These findings reinforce the growing consensus that AI-enabled ECG has the potential to transform cardiovascular screening by enabling earlier and more accessible detection of subclinical cardiac dysfunction.

One of the most significant findings of this review is the superior performance of deep learning models, particularly convolutional neural networks (CNNs), compared to traditional machine learning approaches. Studies such as those by Attia et al. [11,12] and Kwon et al. [13] highlight the ability of CNN-based architectures to extract complex, non-linear features from raw ECG waveforms that are not discernible through conventional interpretation. This capability is especially relevant in the context of LVD, where early electrical changes may be subtle and easily overlooked by clinicians. The high diagnostic accuracy observed across large datasets suggests that AI-ECG can serve as a reliable screening tool, potentially bridging the gap between symptom onset and definitive diagnosis.

Another important contribution of AI-ECG highlighted in this review is its predictive capability. Unlike traditional diagnostic tools that identify established disease, AI models have demonstrated the ability to detect latent patterns associated with future development of LVD. For instance, Raghunath et al. [14] showed that AI-ECG could predict incident left ventricular dysfunction in patients with previously normal echocardiograms. This predictive dimension represents a paradigm shift from reactive to proactive cardiovascular care, enabling early intervention strategies that may delay or prevent progression to overt heart failure. Such an approach aligns with contemporary preventive cardiology frameworks that emphasize risk stratification and early detection.

The review also underscores the expanding applicability of AI-ECG across different ECG modalities. While 12-lead ECG remains the standard and demonstrates the highest diagnostic performance, single-lead and wearable ECG devices have shown promising results [15]. This is particularly relevant in the context of remote monitoring and digital health, where wearable technologies can facilitate continuous, real-time assessment of cardiac function. The integration of AI with wearable ECG devices could enable large-scale population screening and longitudinal monitoring, especially in resource-limited settings where access to echocardiography is constrained. Moreover, studies such as Ribeiro et al. [16] demonstrate the feasibility of deploying AI models on large-scale datasets, further supporting scalability.

Despite these promising findings, several limitations must be considered before widespread clinical implementation. A major concern is the predominance of retrospective study designs, which may introduce selection bias and limit generalizability. Although some studies incorporated external validation cohorts [12,23,24], the overall number remains limited. Models trained on homogeneous datasets may not perform equally well across diverse populations, particularly in low- and middle-income countries where demographic and clinical profiles differ significantly. This highlights the need for multicenter, prospective studies with diverse populations to ensure robustness and equity in AI model performance.

Another critical issue is the “black box” nature of deep learning models, which poses challenges for clinical interpretability and trust. While AI-ECG models can achieve high accuracy, the lack of transparency regarding how predictions are generated may hinder adoption among clinicians. Efforts to improve explainability, such as saliency mapping and feature attribution techniques, are emerging but remain underdeveloped in many studies.

Enhancing interpretability will be essential to facilitate integration into clinical workflows and to meet regulatory requirements.

Additionally, heterogeneity in study methodologies presents challenges in comparing results across studies. Variations in LVEF thresholds (e.g., ≤35%, ≤40%, ≤50%), differences in ECG acquisition protocols, and inconsistencies in performance metrics contribute to variability in reported outcomes. Standardization of definitions and reporting guidelines is necessary to enable meaningful comparisons and meta-analyses in future research.

From a clinical perspective, AI-ECG has significant potential as a screening tool rather than a standalone diagnostic modality. Its high negative predictive value suggests that it could effectively rule out LVD in low-risk populations, thereby reducing unnecessary echocardiographic evaluations. This gatekeeping role could optimize resource utilization, particularly in healthcare systems with limited access to advanced imaging. Furthermore, integration of AI-ECG into routine clinical practice, such as embedding algorithms into ECG machines or electronic health records, could facilitate seamless adoption.

Future research should focus on prospective validation, real-world implementation studies, and cost-effectiveness analyses. Investigating the impact of AI-ECG on clinical outcomes, such as reduction in heart failure incidence or hospitalizations, will be critical to establishing its clinical value. Moreover, combining AI-ECG with other data modalities, such as clinical variables and biomarkers, may further enhance predictive performance and enable more comprehensive risk stratification.

In conclusion, AI-based ECG algorithms represent a promising and rapidly evolving tool for the early detection of left ventricular dysfunction. While current evidence demonstrates high diagnostic accuracy and potential for predictive screening, further validation, standardization, and integration efforts are required to translate these advancements into routine clinical practice.

Conclusion

Artificial intelligence–based ECG algorithms represent a transformative advancement in the early detection of left ventricular dysfunction, demonstrating high diagnostic accuracy and promising predictive capabilities across diverse study populations. By leveraging widely available, low-cost ECG data, these models offer a scalable and accessible approach to screening, with potential to identify subclinical disease and guide timely intervention. However, current evidence is largely derived from retrospective studies with limited external validation, and challenges related to generalizability, interpretability, and standardization persist. Future research should prioritize prospective, multicenter validation and real-world implementation to establish clinical effectiveness and integration into routine care pathways.

References

1. Lee HS, et al. Artificial intelligence–enabled ECG screening for LVSD. 2025.
2. Park S, et al. AI algorithm for early detection of LV dysfunction. 2025.
3. Croon PM, et al. AI-ECG models for LVSD detection. 2026.
4. Velandia H, et al. Systematic review of AI in ECG analysis. 2025.
5. Mayourian J, et al. Deep learning ECG prediction of LV dysfunction. 2025.
6. Bjerkén LV, et al. AI-ECG for LVSD screening: systematic review. 2022.
7. ESC Press Release. AI-ECG early detection of heart failure. 2025.
8. Khunte A, et al. Detection of LV dysfunction using ECG AI. 2023.

9. Zhang S, et al. AI ECG prediction of heart failure. 2026.
10. Gayathiri RR, et al. Diagnostic performance of AI-ECG in LVSD. 2025.
11. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nat Med*. 2019;25(1):70–74.
12. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence–enabled ECG algorithm for the identification of patients with left ventricular systolic dysfunction: a prospective study. *Lancet*. 2022;399(10338):2021–2028.
13. Kwon JM, Kim KH, Jeon KH, et al. Artificial intelligence algorithm for predicting left ventricular systolic dysfunction from ECG. *Eur Heart J*. 2020;41(45):4486–4493.
14. Raghunath S, Pfeifer JM, Ulloa-Cerna AE, et al. Deep neural networks can predict new-onset atrial fibrillation and LV dysfunction from ECG. *Nat Med*. 2021;27(1):112–118.
15. Ko WY, Siontis KC, Attia ZI, et al. Detection of LV dysfunction using a single-lead ECG and AI. *Nat Commun*. 2021;12:1–10.
16. Ribeiro AH, Ribeiro MH, Paixão GMM, et al. Automatic diagnosis of the 12-lead ECG using deep neural networks. *Nat Commun*. 2020;11:1760.
17. Sengupta PP, et al. Cognitive machine-learning algorithm for cardiac imaging and ECG. *Circulation*. 2020;141:180–190.
18. Ouyang D, He B, Ghorbani A, et al. Video-based AI for beat-to-beat cardiac function assessment. *Nature*. 2020;580:252–256.
19. Zhang J, et al. Deep learning ECG model for LV dysfunction detection. *J Am Coll Cardiol*. 2021;77:1189–1199.
20. Cho JH, et al. Artificial intelligence ECG for LVSD detection. *Eur Heart J Digit Health*. 2021;2(3):387–395.
21. Nakajima K, et al. AI-based ECG detection of LV dysfunction. *Heart Vessels*. 2021;36:123–130.
22. Giudicessi JR, et al. AI-enabled ECG for detection of LV dysfunction. *Mayo Clin Proc*. 2021;96:154–165.
23. Kwon JM, et al. Prospective validation of AI-ECG for LVSD screening. *Circulation*. 2023;147:123–135.
24. De Groote P, et al. AI-ECG screening for LV dysfunction in European cohort. *Eur J Heart Fail*. 2022;24:189–198.